國 立 交 通 大 學

統計學研究所

碩 士 論 文

連續時間隨機過程中利用模擬概似函數逼近法改良回歸平均的估計誤差

Estimation Bias for Mean Reversion with Simulated Likelihood Approximation under Continuous Time Stochastic Process

研 究 生 ：周勢耀
指導教授 ：李昭勝 博士

中 華 民 國 九 十 五 年 六 月

連續時間隨機過程中利用模擬概似函數逼近法

改良回歸平均的估計誤差

Estimation Bias for Mean Reversion with Simulated Likelihood Approximation under Continuous Time Stochastic Process

研 究 生：周勢耀　　Student： Shih-yao Zhou

指導教授：李昭勝　　Advisor：Dr. Jack C. Lee

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Statistics
June 2006

Hsin-chu, Taiwan, Republic of China

中華民國九十五年六月

Abstract

This thesis proposes a new method for the estimation of mean reversion effect in diffusion processes from discrete observations. The idea is based on simulating augmented data as high frequency data to cover the inadequacy of discrete observations. The simulation of augmented data is based on Markov-chain Monte Carlo methodology and the estimation of parameters is based on EM algorithm. We implement the Vasicek model as an illustration and the simulation result will be provided. The result demonstrates that the degree of augmentation is quite helpful for the accurate estimation especially when the mean reversion strength is large.

摘要

本論文的目的在於針對離散時間觀察到的擴散過程(Diffusion Process)樣本，提出一個新的方法來估計回歸平均的效果。這構想主要是以模擬出增廣樣本當作高頻資料，以彌補離散時間樣本在估計上的不足。此模擬的程序主要利用馬可夫鏈蒙地卡羅方法(MCMC)，而參數估計則是以 EM 演算法爲基礎。我們以 Vasicek 模型當作例子來測試此方法的可行性。最後可以從模擬結果中發現，當增廣資料的維度提高時，可以在回歸平均強度較大時，得到較好的估計。

誌 謝

在統研所的這兩年裡，尤其我是從政大來到這校風截然不同的交大，一路上要感謝的人實在是數不盡的多。首先是統計所的師長，指導教授李昭勝老師帶領我真正進入統計研究的領域，至於學長牛維方則是給了我論文一個明確的方向，以及吳志強學長時常不厭其煩的和我確認論文每個步驟的細節，可以說沒有他們就沒有我的論文，是我第一個要感謝的。再來就是陪我在研究室度過每一個白天黑夜的損友們，明曄、冠達、火哥、清峰、泓毅、小馬，和你們一起吃飯聊天，日子比較不孤單，也感謝同門的孟樺跟小宛如以及其他同學的互相鼓勵，讓我在最後的論文關鍵期保持高昂的鬥志。

這兩年來除了所上，交大棒球隊更是我重要的回憶之一，我要在此感謝我們教練黃杉楹，他並沒有嫌棄我是從外校來的同學，依然不遺餘力的指導我並給我機會表現，同屆好友永遠的中外野手邱毛、打不穿的浩呆、直球當變化球用的恐龍、永不放棄的施文凱，以及許多努力用心的學弟們跟經理，陪我一起用野球寫下交大生活的另外一頁，在繁忙的研究壓力之餘，有另一個出口。

最後我也要向 G5 的同仁致意，感謝你們的關心以及經驗的傳承，讓我過的很充實，對未來更具信心。

周 勢 耀 僅誌于

國立交通大學統計學研究所

中華民國 95 年 6 月 9 日

Contents

1. Introduction

1.1 Diffusion Process

Diffusion processes have become the standard tool for modeling prices in financial markets for derivative and risk management purposes. Consider an Ito stochastic process that satisfies a stochastic differential equation (SDE) of the form

$$dy(t) = a\{y(t), t, \theta\}dt + b\{y(t), t, \theta\}dW(t) \tag{1.1}$$

where $a\{y(t), t, \theta\}$ and $b\{y(t), t, \theta\}$ are the non-anticipative drift and volatility function respectively, depending on y(t), time t, and an unknown parameter vector θ, and $dW(t)$ is increment of standard Wiener process. Although such continuous time process offer analytic tractability, the parameter that govern their dynamics are often difficult to estimate from discrete time data. In a nutshell, estimation is problematic because the model is formulated in continuous time, while sample data are naturally only available at discrete frequencies. This implies that estimates obtained by naive discretizations of diffusion processes can be subject to discretization bias. Since the direct discretization of diffusion processes will cause estimation bias, the estimation scheme of parameter from discrete observations of y at time-points $0=t_0 < t_1 < \cdots < t_n$, a number of methods have been proposed to estimate diffusion process.

Among all kinds of continuous time model, however, one cannot derive a simple analytic transition density of the process in all cases. If the transition densities $f(y_t \mid y_{t-1}, \theta)$ of y are known, we can use the log- likelihood function

$$L(\theta) = \sum_{t=1}^{n} \log(f(y_t \mid y_{t-1}, \theta)) \tag{1.2}$$

to estimate θ. The corresponding maximum likelihood estimator $\hat{\theta}_n$ is known to have the usual good properties. In the case of time-equidistant observations ($t_i = i\Delta$, $i = 0,1...n$ for some fixed $\Delta > 0$), many papers have proved the consistent and asymptotic normality of $\hat{\theta}_n$ as $n \to \infty$. It is only natural that the number of observations must be large enough for any estimator to be close to the true value, and from a practical point of view it is an important property that ensures the estimator to be close to the true value. Unfortunately,

the transition densities of the continuous time diffusion process are, with the exception of a few special cases, generally unknown or unavailable in closed form so that the conventional likelihood-based inference is often inapplicable.

Traditionally, to overcome the difficulties when the transition densities of y are unknown, the usual alternative is using $\tilde{L}_n(\theta)$ to approximate log-likelihood function based on discrete observations of y. Unless $\max_{1\le i\le n} |t_i - t_{i-1}|$ is "small", nevertheless, in the case of time-equidistant observations, Florens-Zmirou (1989) actually show that estimator based on maximizing $\tilde{L}_n(\theta)$ is consistent. Pederson (1995) derived a sequence $\{l_n^{(N)}(\theta)\}_{N=1}^{\infty}$ of approximations to $L(\theta)$, that gives a connection between $\tilde{L}_n(\theta)$ and $L(\theta)$. The idea is to approximate the (unknown) transition densities $f(y_t \mid y_{t-1}, \theta)$ by a sequence of transition densities $f^{(N)}(y_t \mid y_{t-1}, \theta)$ of approximating Markov process that converge to $f(y_t \mid y_{t-1}, \theta)$ as $N \to \infty$. In the following sections, we will base on this method to discover some estimation problem of special broadly-applied diffusion process in finance.

1.2 Mean Reversion

When using diffusion process as a tool for financial modeling, many researchers may add some components to non-anticipative drift term or volatility function for the purpose of explaining specific phenomenon in financial market, for examples, mean reversion effect or volatility cluster effect. Mean reversion is a tendency for a stochastic process to remain near, or tend to return over time to a long-run average value. Mean reversion effect has been observed in interest rate market, especially in short-term market. In contrast, this behavior is not so obvious in stock market. Vasicek (1977) propose an interest rate model for treasury debt pricing through a mean reversion type stochastic differential equation:

$$dy_t = -\kappa(y_t - \mu)dt + \sigma \cdot dW_t \tag{1.3}$$

where κ is the constant strength of mean reversion, μ is the equilibrium level, σ is the volatility and W_t is the standard Brownian motion.

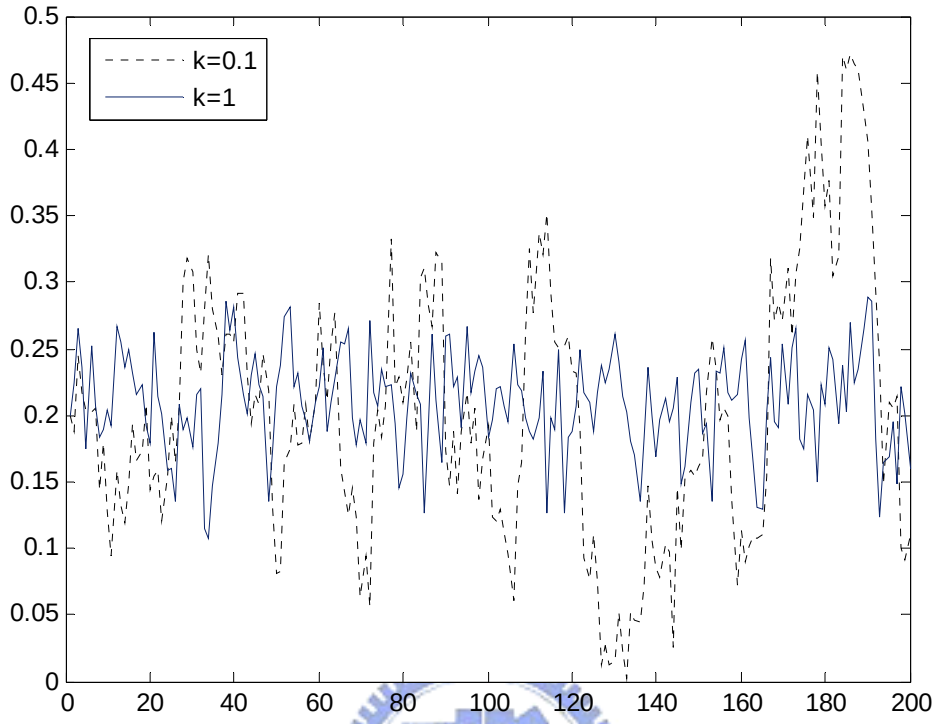


Figure 1: Simulated Data for Comparison of Different Mean Reversion Strength

Such models that incorporate mean reversion effect are also named as mean-reverted process. To be more precise, a process is mean reverted if increments over disjoint intervals are negatively correlated. This is an important property of mean-reverted process because it represents that there is an invisible strength leading the prices back to its equilibrium. Thus asset prices tend to fall (rise) after hitting a maximum (minimum).

1.3 Motivation

Once we had observed such phenomenon in a certain market, a more crucial question comes immediately – how large is the mean reversion strength? That is, even if we can expect the prices will return to the mean, how long does the market need? Many financial practitioners are concerned about this issue. For instance, typically a mean reverted model is used to suggest that an un-hedged long equity position needs less capital than implied in a non-mean reverted model. However, what if we use a mean

reverted model with unknown transition densities? Shall we simply count on discrete observations to make inferences?

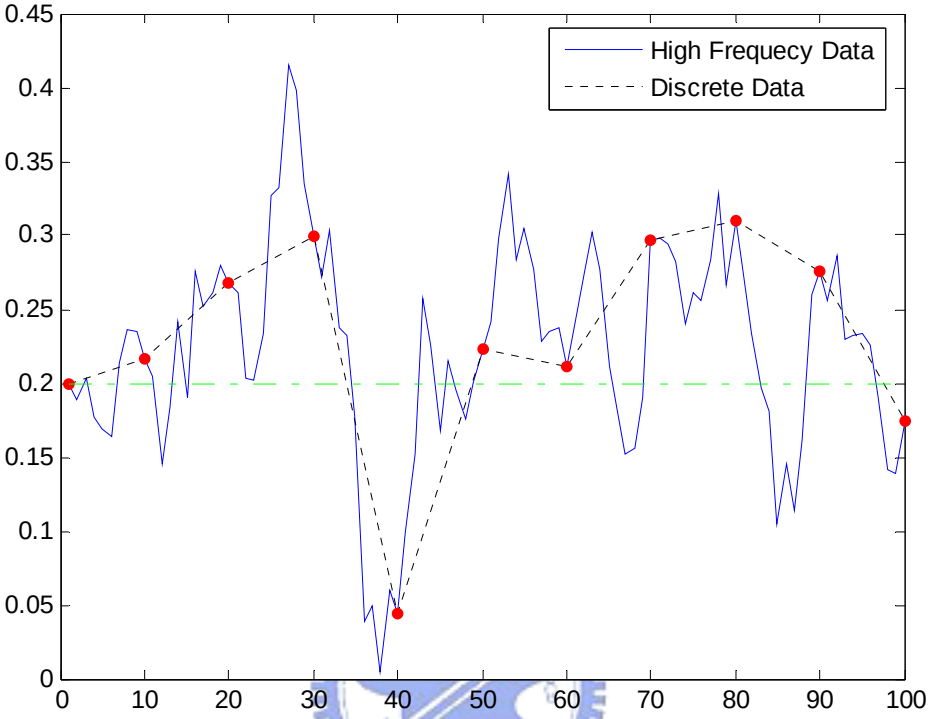


Figure 2: Discretely Observed Data and High Frequency Data

In Figure 2, the dash line and solid line denote 11 discretely observed data and 100 high frequency data respectively, the dash-dot line has the mean of 0.2. If the only information we can obtain is discretely observed data, as we can see in the above figure, this process seems to revert to the mean between the fourth and fifth discrete observations for the first time. Actually, the whole process had reverted to the mean for several times before the discrete observations did. Similar situation also occurs later. With discrete observations barely available, mean reversion strength was under-evaluated in this process. In this thesis, we will show that based on certain process simulating more "augmented data" as high frequency data will improve the estimation merely from discretely observed data. This principle is referred to as "Data Augmentation". The simulating procedure is based on Markov chain Monte Carlo (MCMC) method. The simulated augmented data and discretely observed data together can be called as

complete data so that we can formulate the complete data likelihood function. Simulated likelihood function can be derived after integrating augmented data out of the complete data likelihood function. All the estimation and inference can be conducted through the simulated likelihood function. From the maximum likelihood estimation results, we find that when the degree of augmentation increases, the estimation result can be improved to a certain extent but this depends on the strength of the mean reversion. This corresponds to our claim for the inadequacy of discrete observations for continuous time diffusion process. In fact, the data augmentation is the implementation of the idea of approximate likelihood.

2. Literature Review

2.1 Calibration of Diffusion Process

Based on Ornstein-Uhlenbeck process, Vasicek (1977) proposed a mean reverted process:

$$dy_t = -\kappa(y_t - \mu)dt + \sigma dW_t \tag{1.3}$$

Cox, Ingersoll, Ross (1985), well-known as CIR model, generalized Vasicek model to reflect the effect that volatility changes with the process:

$$dy_t = -\kappa(y_t - \mu)dt + \sigma\sqrt{y_t}dW_t \tag{2.1}$$

These two models are widely-used because the transition densities are unknown. A number of methods have been proposed to estimate diffusion parameters. Working with the difference equation resulting from a Euler discretization of the process can give rise to quasi maximum likelihood or moment-based estimators (e.g., Chan, Karolyi, Longstaff, and Sanders 1992). Naive discretizations, however, give estimates subject to the discretization bias just mentioned if sampling times are infrequent. Yoshida (1992), and Kessler (1997) proposed estimators converging to the true parameter more rapidly as the data are sampled more frequently. An alternative strategy that relies on discretizations of the continuous time likelihood function was given by Liptser and Shiryayev (1977) and Aase (1987), among others. Other analytic methods include those of Sørensen (1995), Bibby and Sørensen (1996) (estimating functions), and Ait-Sahalia (1998) (analytic approximation to the likelihood function), while generalized method of moments (GMM)-based estimators were discussed by Hansen and Scheinkman (1995), and Conley, Hansen, Luttmer, and Scheinkman (1997), among others. Nonparametric methods were proposed by Ait-Sahalia (1996a,b), Jiang and Knight (1997), and Stanton (1997). Previously, simulation-based methods have been proposed for estimating diffusions by the method of simulated moments (Duffie and Singleton 1993), indirect inference methods (Gourieroux, Monfort, and Renault 1993), and the efficient method of moments (EMM) (Gallant and Tauchen 1996), among others. The advantage of simulation-based methods is that they typically apply to more general processes than the analytic methods. For instance, Andersen and Lund (1997) applied EMM in estimation of two- and three-factor nonlinear interest-rate models with unobserved factors.

2.2 Data Augmentation

Pederson (1995) suggested an approximate likelihood function method based on simulating auxiliary variables directly from Euler discretization likelihood function. Elerian, Chib, Shepard (2001) inherited the concepts of approximate likelihood and revised the Euler discretization likelihood simulation to Markov chain Mote Carlo (MCMC) based simulation method. They took CIR model as an illustration to validate the advantages of MCMC simulation method. Eraker (2001) generalized the CIR model to a two-factor constant elasticity of variance model (CEV) with stochastic volatility model. Eraker (2001) believed that MCMC simulations conditioned on more information than Euler discretization simulations recommended by Pederson (1995) because MCMC method conditioned on the whole observations and interpolates a certain number of augmented data into each pairs of observations. Both Elerian, Chib, Shepard (2001) and Eraker (2001) used Bayesian approaches to estimate diffusion parameters but the former used informative priors and the other exploited non-informative priors. Niu and Lee (2006) proposed the GARCH based simulated likelihood approximations for continuous time stochastic volatility models and applied this method to option pricing.

3. Methodology

In this paper we try to validate the advantage of augmented variables by declaring that the approximate log-likelihood function will converge to the transition densities as the degree of augmentation increases. By taking augmented variables as the high-frequency data, we can improve the calibration of model because the incorporation of high-frequency data together with observed data are denser in some specific time interval. Especially for the strength of mean reversion κ, it is usually difficult to observe its level in a system with big fluctuation. When strength of mean reversion κ is weak, the path of observed data tends to be smoother and requires a longer period to return to its equilibrium. In contrast, when κ is strong, the path may exhibit more feature and fast mean-reverted, so it may be naïve if we merely use low frequency data to catch the structure. For this reason, we first assume a continuous time model and simulate augmented variables which follow this model. However, it is difficult to handle the case with unknown transition densities. As a result, we use a simple Ornstein-Ulenbeck process, modified by Vasicek in 1997 with the incorporation of mean reversion effect since the close form solution can be easily obtained. Here we simulate observed data from Vasicek model instead of using real time data because the true dynamics of data are unobservable.

We assume that asset dynamics follow equation (1.3), which is the well-known Vasicek interest rate model. To begin with, consider the Euler approximation of the SDE

$$y_{t+1} - y_t = -\kappa(y_t - \mu)\Delta + \sigma(W_{t+1} - W_t) \tag{3.1}$$

, under which the transitional density is

$$f(y_{t+1} \mid y_t) = \phi(y_{t+1} \mid y_t - \kappa(y_t - \theta)\Delta, \sigma^2\Delta) \tag{3.2}$$

, where $\phi(\cdot \mid a, b)$ denotes the normal density with mean a and variance b. Although this is the simplest discrete time approximation of the SDE, however, it is normally too coarse to approximate the true transition density adequately. Hence we want to propose an improved method through the utilization of auxiliary variables together with the discretely observed data to approximate the continuous time financial model.

Here we shall use the famous Markov Chain Monte Carlo method to simulate auxiliary

variables and the elaborate EM algorithm to estimate parameters. Now let us denote Y^* as the auxiliary variables we simulated in each sub-interval of observed data *Y*. Thus, our simulation procedure in general can be summarized as follows:

General sampling scheme

1. Initialize Y*, θ.
2. Update y_t^* from $y_t^* \mid y_t, y_{t+1}, \theta$, for t=1,2,…,T-1
3. Update θ from $\theta \mid Y^*, Y$.
4. Record the value of θ and go to step 2. Repeat for a large number of times.

3.1 MCMC method

In recent years statisticians have been increasingly drawn to Markov chain Monte Carlo (MCMC) methods to simulate complex, nonstandard multivariate distribution. The Gibbs sampling algorithm is one of the best known of the methods (see Casella and George(1992)). The Gibbs sampler is a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density. In this paper we mainly use the Gibbs sampler to simulate the auxiliary variables instead of Metropolis-Hasting (M-H) algorithm since the conditional distribution of auxiliary variables on discretely observed data can be formulated in a simple close form. Through the use of techniques like Gibbs sampler, we are able to avoid difficult calculations, replacing them with a sequence of easier calculations.

Markov chain Monte Carlo sampling from θ, $Y^* \mid Y$ is achieved by sampling in turn the full conditional distributions $Y^* \mid Y, \theta$ and $\theta \mid Y^*, Y$. One iteration of the Markov chain is completed by revising both Y* and θ from these two distributions. A simple calculation (based on the Markov property of the diffusion) show that the full conditional distribution can be expressed as

$$f(Y^*, Y \mid y_1, \theta) = \prod_{t=1}^{T-1} \prod_{j=1}^{M+1} f(y_t^* \mid y_{t-1}^*, \theta) \tag{3.3}$$

due to the fact that the augmented data y_t^* is conditionally independent of the remaining

augmented data, given (y_t, y_{t+1}, θ). This is done by generating a "Gibbs sequence" of random variables:

$$(y^*_{t,1}, y^*_{t,2}, \cdots, y^*_{t,M})_{(1)}, \quad t = 1, 2, \cdots, T-1$$

.

.

.

$$(y^*_{t,1}, y^*_{t,2}, \cdots, y^*_{t,M})_{(K)}, \quad t = 1, 2, \cdots, T-1 \tag{3.4}$$

Note that the initial values $(y^*_{t,1}, y^*_{t,2}, \cdots, y^*_{t,M})_{(1)}$ need to be specified, the rest of all is obtained iteratively by alternatively generating values as

$$(y^*_{t,1}, y^*_{t,2}, \cdots, y^*_{t,M})_{(2)} \sim f(y^*_{t\,(2)} \mid y_t, y_{t+1}, y^*_{t\,(1)}; \theta)$$

.

.

$$(y^*_{t,1}, y^*_{t,2}, \cdots, y^*_{t,M})_{(K)} \sim f(y^*_{t\,(K)} \mid y_t, y_{t+1}, y^*_{t\,(K-1)}; \theta) \tag{3.5}$$

In short, at the i-th iteration of Gibbs sampler, we draw

$$y^{*(i)}_{t,j} \sim f(y^*_{t,j} \mid y^{*(i)}_{t,j-1}, y^{*(i-1)}_{t,j+1}; \theta) \tag{3.6}$$

where j=1,2,...,M. Since the information $y^{*(i)}_{t,j+1}$ is not available at the current iteration, we should condition on the last iteration result $y^{*(i-1)}_{t,j+1}$.

We refer to this generation as Gibbs sampling. It turns out that under reasonably general conditions, the distribution of Y^* converges to the approximate likelihood functions.

3.2 Example

Vasicek model can be solved explicitly and represented as

$$y_t \sim N(y_{t-1}e^{-\kappa} + \mu(1-e^{-\kappa}), \frac{\sigma^2(1-e^{-2\kappa})}{2\kappa}) \tag{3.7}$$

By the above equation we can obtain the true transition probability $g(y_{t+1} \mid y_t)$.

Now consider the detail of simulating auxiliary data Y* between each pair of discretely observed data Y, conditioning on these discretely observed data we can obtain

$$f(Y^* \mid Y, \Theta) = \prod_{t=1}^{T-1} f(y_t^* \mid y_t, y_{t+1}, \theta) \tag{3.8}$$

where y_t^* = ($y_{t,1}^*, y_{t,2}^*, \ldots\ldots\ldots\ldots, y_{t,M}^*$), and $f(y_t^* \mid y_t, y_{t+1}, \theta)$ is the density from Euler approximation.

Since $f(y_t \mid y_{t-1}, y_{t+1}, \theta)$ is proportional to $f(y_{t+1} \mid y_t, \theta) f(y_t \mid y_{t-1}, \theta)$, thus we can derive the conditional distribution

$$y_t \mid y_{t-1}, y_{t+1}, \theta \sim N(\frac{1}{1+\rho^2}[\rho(y_{t-1} + y_{t+1}) + \kappa^2\Delta^2\mu], \frac{\sigma^2\Delta}{1+\rho^2}) \tag{3.9}$$

,where $\rho = 1 - \kappa\Delta$, t=1,2,……,T-1

If we simulate M equally-spaced auxiliary variables in each period and interpolate those into the proper interval of discretely observed data, then the available size of data can be augmented from T to (T-1)(M+1) + 1. Now let $y_{t,j}^*$ denote the j-th auxiliary variable in the t-th interval of discretely observed data, as an analogy, the transition probability can be written as

$$y_{t,j}^* \mid y_{t,j-1}^*, y_{t,j+1}^*, \theta \sim N(\frac{1}{1+\rho^2}[\rho(y_{t,j-1}^* + y_{t,j+1}^*) + \kappa^2(\Delta^*)^2\mu], \frac{\sigma^2\Delta^*}{1+\rho^2}) \tag{3.10}$$

t=1,2,……,T-1, j=1,2,…….,M, and note that $y_{t,0}^* = y_t$, $y_{t,M+1}^* = y_{t+1}$, $\Delta^* = \frac{\Delta}{M+1}$.

This sampling procedure should be conducted through Gibbs sampler because for each

$y_{t,j}^{*}$ conditioned on $y^{*}_{t,j-1}$ and $y^{*}_{t,j+1}$, when $y^{*}_{t,j+1}$ is still unknown so that we need to guess initial value to run the algorithm. If the algorithm is implemented for only one time, such a rough method may cause a fatal error. As a result, it is better to iterate the algorithm for a large number of times until convergence.

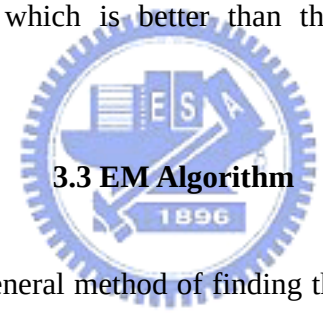
With the utilization of auxiliary variables, the transition density under Euler approximation is

$$f(y^{*}_{t,j} \mid y^{*}_{t,j-1}, \theta) = \phi(y^{*}_{t,j} \mid y^{*}_{t,j-1} - \kappa(y^{*}_{t,j-1} - \mu)\Delta^{*}, \sigma^{2} \cdot \Delta)\,. \tag{3.11}$$

So the complete data likelihood function can be formulated as

$$f(Y, Y^{*} \mid y_1, \theta) = \prod_{t=1}^{T-1} \prod_{j=1}^{M+1} f(y^{*}_{t,j} \mid y^{*}_{t,j-1}, \theta) \tag{3.12}$$

Based on the complete data likelihood function, we can obtain an improved maximum likelihood estimator (MLE), which is better than the one derived from the Euler approximation of SDE.



3.3 EM Algorithm

The EM algorithm is a general method of finding the maximum likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing value. There are two main applications of the EM algorithm. The first occurs when the data indeed has missing values, due to problems with limitations of the observation process. The second occurs when optimizing the likelihood function can be simplified by assuming the existence of additional but missing data. The later application is more common in the computational pattern recognition community. In our case, we assume the data Y is discretely observed and is generated from some distribution. We call Y the incomplete data and Y* (auxiliary variables we simulated) the missing data. We assume that a complete data set Z = (Y, Y*) and also assume a joint density function:

$$p(z \mid \Theta) = p(Y, Y^{*} \mid \Theta) = p(Y^{*} \mid Y, \Theta)\, p(Y \mid \Theta) \tag{3.13}$$

With this new density function, we can define a new likelihood function,

$L(\Theta \mid Z) = p(Y, Y^* \mid \Theta)$, called the complete-data likelihood. Note that this function is in fact a random variable since the missing information Y* is unknown and presumably governed by an underlying distribution. The original likelihood function $L(\Theta \mid Y)$ is referred to as the incomplete-data likelihood function.

The EM algorithm first finds the expected value of the complete-data log-likelihood $\log p(Y, Y^* \mid \Theta)$ with respect to the unknown data Y* given the observed data Y and the current parameter estimates. That is, we define:

$$Q(\Theta, \Theta^{(i-1)}) = E[\log P(Y, Y^* \mid \Theta) \mid Y, \Theta^{(i-1)}] \tag{3.14}$$

where $\Theta^{(i-1)}$ are the current parameters estimates that we used to evaluate the expectation and Θ are the new parameters that we optimize to increase Q. The key element to understand is that Y and $\Theta^{(i-1)}$ are constants, Θ is a normal variable that we wish to adjust, and Y* is a random variable governed by the distribution $f(y^* \mid Y, \Theta^{i-1})$. The right side of the above equation can be represented as

$$E[\log P(Y, Y^* \mid \Theta) \mid Y, \Theta^{(i-1)}] = \int_{y^*} \log p(y, y^* \mid \Theta) f(y^* \mid y, \Theta^{(i-1)}) dy^* \tag{3.15}$$

Note that $f(y^* \mid Y, \Theta^{i-1})$ is the marginal distribution of the unobserved data Y* and is dependent on both the observed data Y and $\Theta^{(i-1)}$. In the best of cases, this marginal distribution is a simple analytical expression of the assumed parameters $\Theta^{(i-1)}$ and perhaps the data. In the worst of the cases, this density might be very hard to obtain due to the high dimensionality of augmented variables. So we must use the numerical method to approximate the true expectation. Our approach is based on the idea of Monte Carlo integration, i.e., to simulation R identically and independently distributed paths of augmented data conditioned on discretely observed data.

$$E[\log P(Y, Y^* \mid \Theta) \mid Y, \Theta^{(i-1)}] \approx \frac{1}{R} \sum_{r=1}^{R} \log(\prod_{t=1}^{T-1} \prod_{j=1}^{M+1} f(y_{t,j}^{*(r)} \mid y_{t,j-1}^{*(r)}, \Theta) \tag{3.16}$$

Note that $y_{t,j}^{*(r)}$ is one of the augmented data from r-th path conditioned on $\Theta^{(i-1)}$ and Y.

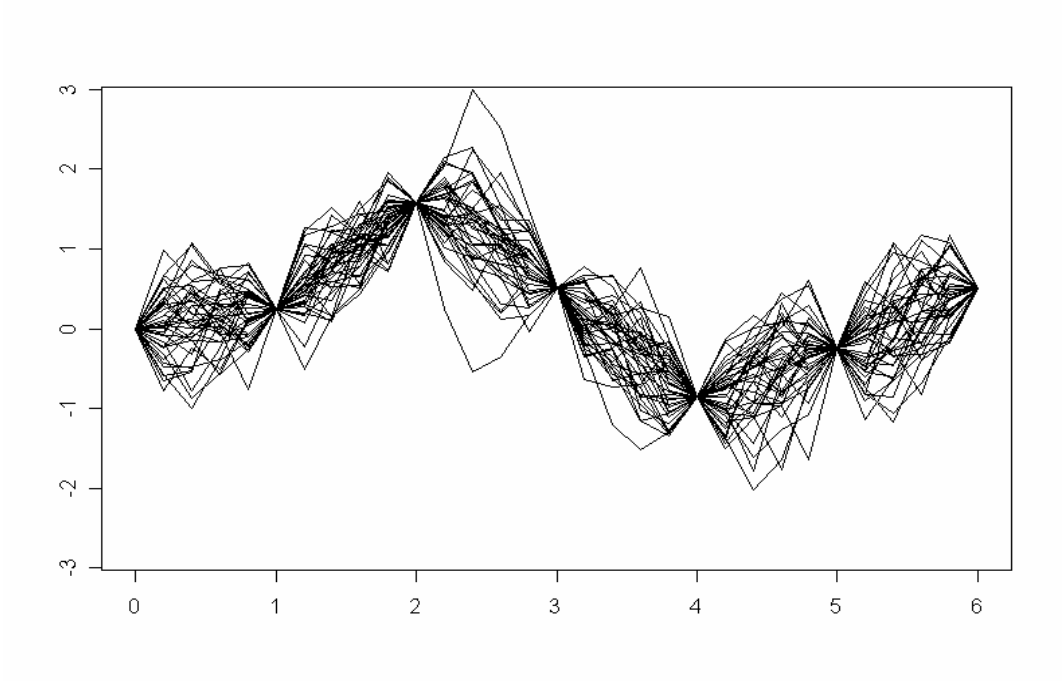


Figure 3: **7 Observed Data and 20 Simulated Paths**

The evaluation of this expectation is called the E-step of the algorithm. Notice the meaning of the two arguments in the function $Q(\Theta, \Theta^{(i-1)})$ .The first argument Θ corresponds to the parameters that ultimately will be optimized in an attempt to maximize the likelihood function. The second argument $\Theta^{(i-1)}$ corresponds to the parameters that we use to evaluate the expectation.

The second step of the EM algorithm, also called M-step, is to maximize the expectation we computed in the first step. That is, we find:

$$\Theta^{(i)} = \arg\max_{\Theta} Q(\Theta, \Theta^{i-1}) \tag{3.17}$$

These two steps are repeated as necessary. The advantage of the EM algorithm guaranteed that each iteration increases log-likelihood and will finally converge to a local maximum of the log-likelihood function. There are many rate-of-convergence papers but we will not discuss this issue here. A complete procedure to get a convergent maximum likelihood estimate is to simulate R independent paths conditioned on $\Theta^{(0)}$ and then numerically search for a steady estimate. Conditioned on the new estimate $\Theta^{(1)}$, simulating R independent paths again to acquire a new estimate $\Theta^{(2)}$. Repeat these steps

until estimates converge.

As presented above, we basically introduce the EM algorithm and so the algorithm is presented in its most general form. So far, it is still not clear how we "code up" the EM algorithm applied to our case. Now, based on the transition probability of Vasicek model, we propose our estimation algorithm as follows:

1. For i =1, conditioned on $\Theta^{(i-1)}$ and discretely observed data Y, simulate Y* by MCMC method.
2. (E-step)

   $Q(\Theta,\Theta^{(i-1)}) = E[\log P(Y,Y^* \mid \Theta) \mid Y,\Theta^{(i-1)}]$, here the close form for $Q(\Theta,\Theta^{(i-1)})$ is difficult to obtain, therefore we need the help of Monte Carlo approximation. Let $Y^{*(r)}$ denote r-th path of latent data, r=1, 2,….., R.

   As R is large, $Q(\Theta,\Theta^{(i-1)})$ =

$$\frac{1}{R}\sum_{r=1}^{R}\log(Y,Y^{*(r)} \mid \Theta) = \frac{1}{R}\sum_{r=1}^{R}\log(\prod_{t=1}^{T-1}\prod_{j=1}^{M+1} f(y_{t,j}^{*(r)} \mid y_{t,j-1}^{*(r)},\Theta)\,. \qquad (3.18)$$

   Note that $y_{t,j}^{*(r)}$ is sampled from $f(y_{t,j}^{*(r)} \mid y_{t,j-1}^{*(r)}, y_{t,j+1}^{*(r)},\Theta^{(i-1)})$.

3. (M-step)

$$\text{Let } \Theta^{(i)} = \underset{\Theta}{Max}\, Q(\Theta,\Theta^{(i-1)}) = \underset{\Theta}{Max}\frac{1}{R}\sum_{r=1}^{R}\log(Y,Y^{*(r)} \mid \Theta) \qquad (3.19)$$

4. i = i+1, and then go to step 1.

4. Simulation Result

In this section, we present estimates of the parameters in the Vasicek model with different κ. M denotes the degree of augmentation, which means the number of augmented data interpolated into each pair of observed data. The discretely observed data length is 200. Here we uses M=1, 2, 4, 6, 8, and κ=0.1, 0.5, 1. μ and σ in 3 cases of κ are fixed at 0.2 and 0.05 respectively. Our main goal is to validate that different κ also have different sensitivities to the degree of augmentation. For the sake of validating our method, we first simulate 100 sets of observed data from Vasicek model as the observed data instead of real time data. For each set we simulate large numbers of augmented data and do simulated maximum likelihood estimation to obtain one set of estimate (κ, μ, σ). Note that when implementing EM algorithm, we take the maximum likelihood estimation from true Vasicek transition densities as the initial value. Hence there are 100 sets of parameter estimates which can be used to calculate mean, bias, standard deviation (SD), mean square error (MSE), and mean absolute relative error (MARE).

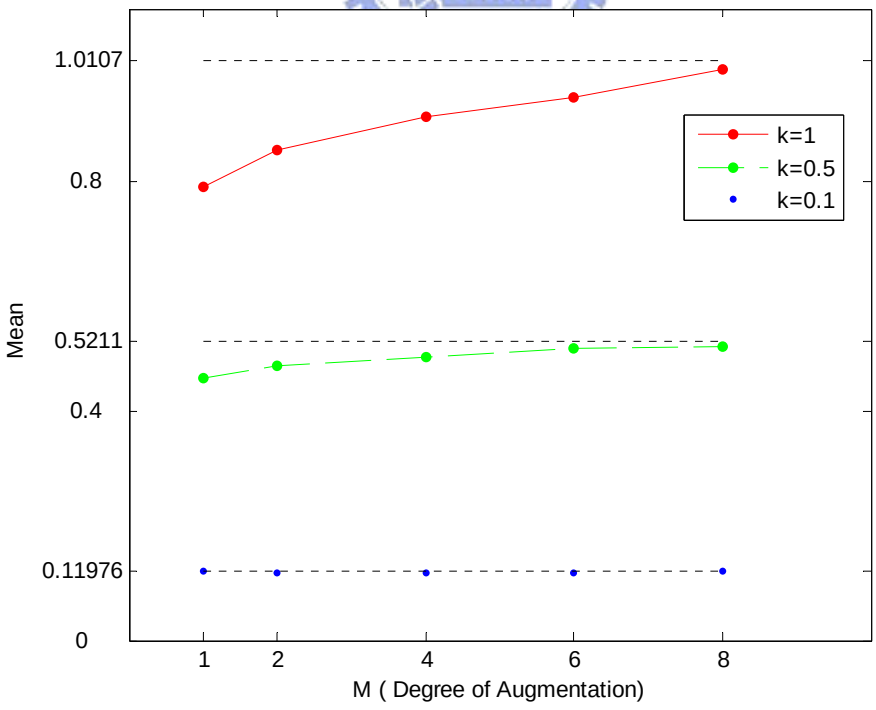


Figure 4: Mean of kappa Estimates

In Figure 4, we can see that when κ =0.1, the advantage of utilizing data augmentation is insignificant. While in the case of κ =0.5, the estimates κ have begun to slowly improve. In the largest case of κ =1, the estimates converge to the true value after a certain number of augmented data were interpolated. When M=1, the estimates have strong biases because the mean only attained to 0.78. This corresponds to our initial guess that infrequently observed data will give rise to discretization bias, especially when the mean reversion strength κ is large. The dash line is the result of maximum likelihood estimation from the true transition densities of Vasicek model. All the estimation results from data augmentation will eventually converge to the maximum likelihood estimation from the true transition densities.

From the view points of mean square error (MSE), Figure 5 also provides a persuasive result. When κ =1, MSE declines fast as the degree of augmentation increases.

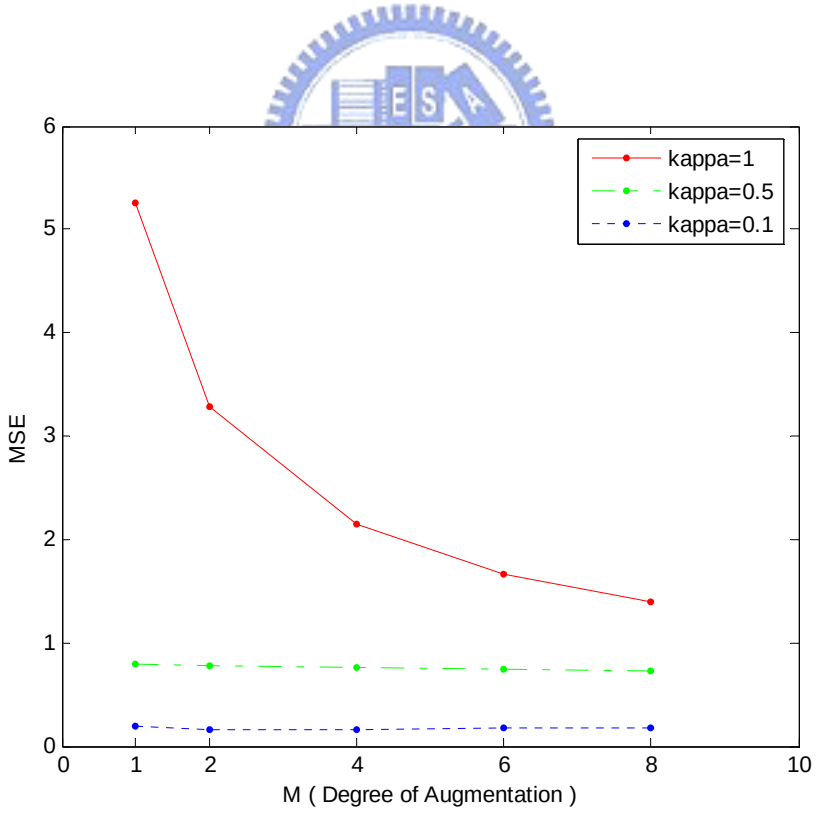


Figure 5: Mean Square Error of kappa and note that the result is MSE times 100

The last indicator is the percentage mean absolute error (MARE). One may suspect that large κ leads to a large magnitude of revision in value. If we do not check the MARE we may probably over-evaluate the effectiveness of data augmentation method. In Figure 6, MARE will indicate that this result is not a coincidence.

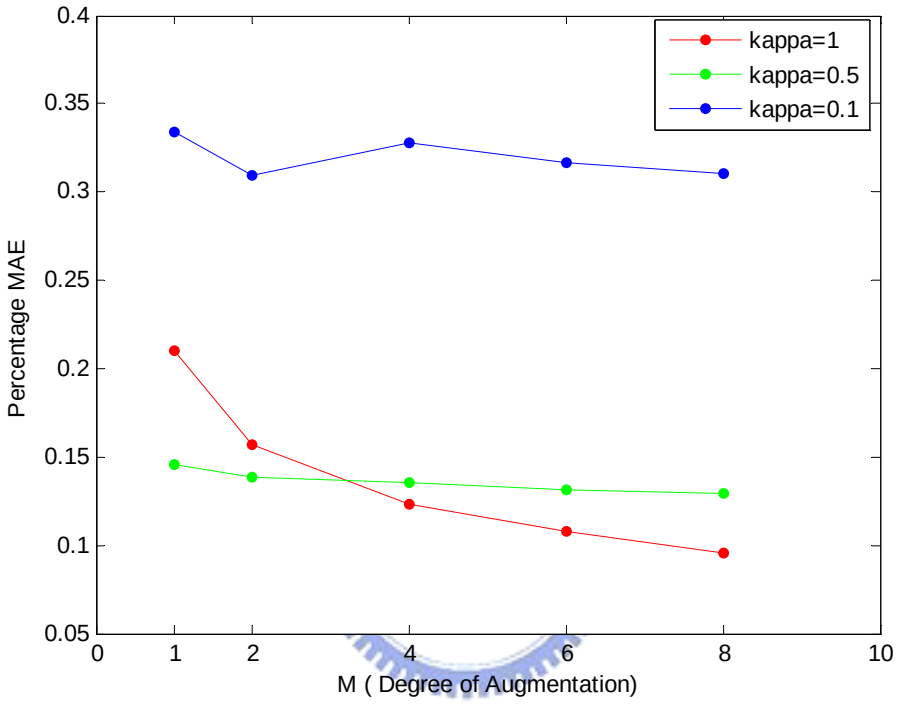


Figure 6: Mean Absolute Relative Error Plot, MARE = $\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{\kappa}_i^{(M)}-\kappa}{\kappa}\right|$**, N=100**

Obviously, when κ =0.1 the MARE is larger than another two cases uniformly. This is a powerful evidence that the improvement of data augmentation at κ =1 and κ =0.5 is not from its larger magnitude in value but truly from its bias. When $M \geq 4$, the MARE of κ =1 is smaller than that of κ =0.5, hence in this measurement we clearly see that data augmentation method is quite helpful for large κ cases.

The following three tables list all the estimates of (κ, μ, σ). Amazingly, among all three parameters κ is the most sensitive one to the degree of augmentation. While μ is the least one, actually numerous literature have shown that the estimation of drift term seldom cause a big trouble.

Simulation Result (κ =0.1)

	Mean	Bias	s.d.	MSE*N	MARE
(κ =0.1)					
M=1	0.1218	0.0218	0.0382	0.1903	0.3339
M=2	0.1170	0.0170	0.0364	0.1588	0.3093
M=4	0.1171	0.0170	0.0365	0.1596	0.3281
M=6	0.1181	0.0181	0.0370	0.1671	0.3163
M=8	0.1208	0.0208	0.0377	0.1827	0.3104
MLE	0.1197	0.0197	0.0382	0.1822	0.3280
(μ =0.2)					
M=1	0.1993	-0.0006	0.0325	0.1036	0.1237
M=2	0.1993	-0.0006	0.0325	0.1036	0.1236
M=4	0.1993	-0.0006	0.0325	0.1035	0.1236
M=6	0.1993	-0.0006	0.0325	0.1037	0.1237
M=8	0.1993	-0.0006	0.0325	0.1036	0.1236
MLE	0.1993	-0.0006	0.0325	0.1036	0.1236
(σ =0.05)					
M=1	0.0489	-0.0010	0.0024	0.0007	0.0319
M=2	0.0489	-0.0010	0.0024	0.0007	0.0429
M=4	0.0493	-0.0006	0.0023	0.0006	0.0340
M=6	0.0500	0.0000	0.0021	0.0004	0.0385
M=8	0.0502	0.0002	0.0020	0.0004	0.0426
MLE	0.0500	0.0000	0.0025	0.0006	0.0382

Table 1: The simulation result is based on Vasicek model. N=100 sets of data, each set has length of discrete observations T=200, and iterates 50 times for Gibbs sampling to be stationary, and R=100 independent paths to approximate log-likelihood expectation. MSE*N represents $\sum_{i=1}^{N}(\hat{\theta}_i - \theta_{true})^2$, *N*=100.

Simulation Result (κ =0.5)

	Mean	Bias	s.d.	MSE*N	MARE
(κ =0.5)					
M=1	0.4564	-0.0436	0.0788	0.7971	0.1454
M=2	0.4768	-0.0231	0.0858	0.7761	0.1390
M=4	0.4931	-0.0068	0.0882	0.7684	0.1350
M=6	0.5084	0.0084	0.0871	0.7513	0.1313
M=8	0.5129	0.0129	0.0855	0.7343	0.1294
MLE	0.5211	0.0210	0.1046	0.8692	0.1452
(μ =0.2)					
M=1	0.1989	-0.0010	0.0077	0.0059	0.0306
M=2	0.1989	-0.0010	0.0077	0.0060	0.0307
M=4	0.1989	-0.0010	0.0077	0.0060	0.0306
M=6	0.1989	-0.0010	0.0077	0.0060	0.0306
M=8	0.1989	-0.0010	0.0077	0.0060	0.0307
MLE	0.1988	-0.0011	0.0077	0.0060	0.0309
(σ =0.05)					
M=1	0.0442	-0.0057	0.0026	0.0039	0.1164
M=2	0.0460	-0.0039	0.0028	0.0023	0.0841
M=4	0.0476	-0.0023	0.0028	0.0013	0.0617
M=6	0.0487	-0.0012	0.0026	0.0008	0.0472
M=8	0.0492	-0.0007	0.0024	0.0006	0.0410
MLE	0.0503	0.0003	0.0037	0.0012	0.0558

Table 2: The simulation result is based on Vasicek model. 100 sets of data, each set has length of discrete observations T=200, and iterates 50 times for Gibbs sampling to be stationary, and R=100 independent paths to approximate log-likelihood expectation. MSE*N represents $\sum_{i=1}^{N}(\hat{\theta}_i - \theta_{true})^2$, *N*=100.

Simulation Result (κ =1)

	Mean	Bias	s.d.	MSE*N	MARE
(κ =1)					
M=1	0.7896	-0.2104	0.0947	5.2604	0.2103
M=2	0.8541	-0.1458	0.1094	3.2807	0.1566
M=4	0.9126	-0.0873	0.1188	2.1393	0.1228
M=6	0.9466	-0.0534	0.1187	1.6625	0.1078
M=8	0.9939	-0.0061	0.1191	1.3940	0.0953
MLE	1.0107	0.0106	0.1592	2.4960	0.1290
(μ =0.2)					
M=1	0.2003	0.0003	0.0037	0.0013	0.0143
M=2	0.2002	0.0002	0.0037	0.0013	0.0144
M=4	0.2002	0.0002	0.0037	0.0013	0.0142
M=6	0.2002	0.0002	0.0037	0.0013	0.0143
M=8	0.2002	0.0002	0.0037	0.0013	0.0143
MLE	0.2002	0.0002	0.0037	0.0013	0.0143
(σ =0.05)					
M=1	0.0397	-0.0102	0.0021	0.0107	0.2045
M=2	0.0427	-0.0072	0.0023	0.0057	0.1452
M=4	0.0455	-0.0045	0.0022	0.0025	0.0914
M=6	0.0469	-0.0030	0.0020	0.0013	0.0634
M=8	0.0487	-0.0012	0.0018	0.0004	0.0353
MLE	0.0503	0.0003	0.0035	0.0012	0.0560

Table 3: The simulation result is based on Vasicek model. 100 sets of data, each set has length of discrete observations T=200, and iterates 50 times for Gibbs sampling to be stationary, and R=100 independent paths to approximate log-likelihood expectation. MSE*N represents $\sum_{i=1}^{N}(\hat{\theta}_i - \theta_{true})^2$, *N*=100.

Simulation Results

	κ =0.1	κ =0.5	κ =1
(Mean)			
M=1	0.1218	0.4564	0.7896
M=2	0.1170	0.4768	0.8541
M=4	0.1171	0.4931	0.9126
M=6	0.1181	0.5084	0.9466
M=8	0.1208	0.5129	0.9939
MLE	0.1197	0.5211	1.0107
(MSE*N)			
M=1	0.1903	0.7971	5.2604
M=2	0.1588	0.7761	3.2807
M=4	0.1596	0.7684	2.1393
M=6	0.1671	0.7513	1.6625
M=8	0.1827	0.7343	1.3940
MLE	0.1822	0.8692	2.4960
(MARE)			
M=1	0.1454	0.1454	0.2045
M=2	0.1390	0.1390	0.1452
M=4	0.1350	0.1350	0.0914
M=6	0.1313	0.1313	0.0634
M=8	0.1294	0.1294	0.0353
MLE	0.1452	0.1452	0.0560

Table 4: Comparison of Mean, MSE, MARE at different κ

5. Conclusions

So far, unlike mean and volatility, there is no existing universal measure of mean reversion. Investment professionals have observed this phenomenon but still could not develop an effective way to quantify it. Our effort in this thesis does not represent that we have developed a simple method to measure mean reversion. In past years many people had tried but this issue still remained inconclusive. What we want to provide is an effective and reliable procedure to estimate mean reversion strength. Though we only take the Vasicek model as an illustration, in fact this method can be applied to all models without analytic transition densities. Especially, the data augmentation method deals with a frequently-ignored question – different difficulties in the estimation while mean reversion strength is big or small. In the simulation result we have shown that when mean reversion speed is slow the estimation result does not improve much as the degree of data augmentation increases. However if mean reversion speed is fast, the estimation result greatly improves as the degree of data augmentation increases. Meanwhile, the entire procedure was based on standard statistical tools, like MCMC sampling and EM algorithm for maximum likelihood estimation. Thus the canonical statistical inference can be applied.

References

Ait-Sahalia, Y. (1996a), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

--- (1996b), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studie*, 9, 385–426.

--- (1998), "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach," working paper, Princeton University, Dept. of Economics.

Andersen, T. G., and Lund, J. (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short-term Interest Rate," *Journal of Econometrics*, 7, 343–377.

Bibby, B. M., and Sørensen, M. (1996), "On Estimation for Discretely Observed Diffusions: A Review," *Theory of Stochastic Processes*, 2, 49–56.

Bilmes,J.A. (1998), "A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," unpublished manuscript, International Computer Science Institute Berkeley CA, 94704 and Computer Science Division Department of Electrical Engineering and Computer Science U.C. Berkeley TR-97-021.

Casella, G., George,E.I., "Explaining the Gibbs Sampler," *The American Statistician*, Vol. 46, No. 3. (Aug., 1992), pp. 167-174.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Conley, T., Hansen, L. P., Luttmer, E., and Scheinkman, J. (1997), "Estimating Subordinated Diffusions from Discrete Time Data," *Review of Financial Studies*, 10, 525–577.

Duffie, D., and Glynn, P. (1996), "Estimation of Continuous-Time Markov Process Sampled at Sampled at Random Time Intervals," unpublished manuscript, Stanford University, Department of Economics.

Eraker, B., (2001), "MCMC Analysis of Diffusion Models with Application to Finance," *Journal of Business & Economic Statistics*, April 2001, Vol. 19, No 2.

Elerian, O., Chib, S., and Shephard, N. (1999), "Likelihood Inference of Discretely

Observed Nonlinear Diffusions," *Econometrica*, Vol. 69, No.4 (July, 2001), 959-993.

Gallant, A. R., and Tauchen, G. (1996), "Which Moments to Match?" *Econometric Theory*, 12, 657–681.

Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Economics*, 8, S85-S118.

Hansen, L. P., and Scheinkman, J. (1995), "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804. Stanford University, Dept. of Economics.

Jiang, G. J., and Knight, J. L., (1997), "A Nonparametric Approach to the Estimation of Diffusion Processes, With an Application to a Short-term Interest Rate Model," *Econometric Theory*, 13, 615–645.

Kessler, M. (1997), "Estimation of an Ergodic Diffusion From Discrete Observations," Scandinavian Journal of Statistics, 24, 1-19.

Liptser, R. S., and Shiryayev, A. N. (1977), Statistics of Random Processes I: General Theory, New York: Springer-Verlag.

Niu, W.F., and Lee, J.C. (2006), "GARCH Based Simulated Likelihood Approximations for Continuous Time Stochastic Volatility Models," Working Paper, National Chiao Tung University, Hsinchu, Taiwan.

Sørensen, M. (1995), "Estimating Functions for Discretely Observed Diffusions: A Review," Research Report 348, University of Aarhus, Dept. of Mathematics.

Pedersen, A. (1995), "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973–2002.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Yosida, N. (1992), "Estimation for Diffusion Processes From Discrete Observations," *Journal of Multivariate Analysis*, 41, 220-242.